EMMA YU

emmayu2918@gmail.com   (216) 313-2443

Portfolio: https://www.emmajyu.com   GitHub: https://github.com/emmayu18/

EDUCATION

Northwestern University   *Graduation: June 2023*

Evanston, IL

Bachelor of Arts in **Data Science** & **Neuroscience** – GPA: 3.73

- Relevant coursework: Data Science Project, Data Visualization, Advanced Machine Learning, Introduction to Artificial Intelligence, Data Management and Information Processing, Statistical Theory and Methods

EXPERIENCE

Oncology Research Assistant   *June 2020 – Present*

Northwestern University Feinberg School of Medicine – Chicago, IL

- Led a team of 6 undergraduate peers in an AI/ML genetic variant classification tool evaluation project
- Formulated the data collection methodology and oversaw the delegation of the collection process
- Published an abstract in the American Association for Cancer Research and submitted a manuscript to journals
- Presented a poster for the AI classification tool evaluation project at the AACR Annual Meeting 2022
- Instructed team members on basic machine learning concepts and data analysis

Data Science Research Assistant   *June 2023 – August 2023*

Brigham and Women's Hospital – Boston, MA

- Reverse engineered a Python code for a BERT neural network model to develop a user-friendly predictor script
- Introduced the predictor script and documentation to a client and provided comprehensive training on its usage
- Experimented with topic segmentation to improve the performance of a natural language processing tool designed to extract unstructured data from medical charting systems
- Inspected through over 500 model errors to locate areas that could be targeted by topic segmentation

Data Science Intern   *June 2022 – August 2022*

Intelligent Medical Objects (IMO) – Rosemont, IL

- Led a classification model-building project to identify junk terms from medical terminology processing workflow
- Collaborated with a senior data scientist and gained experience working with real-world data and advanced machine learning techniques
- Engineered extensive features using NLP and achieved a model performance metric F1-score of 0.98
- Presented promising results to the Director of Content Delivery and received confirmation to take the model to production

COURSES/PROJECTS

Data Science Project Course   *March 2023 – June 2023*

Northwestern University – Evanston, IL   https://2r7ec0-emma-yu.shinyapps.io/mcmf_map/

- Created a user-friendly interactive data visualization app to address the City of Chicago Office of the Mayor's request for a measure of equity in afterschool programs across different age groups and regions in Chicago
- Built a stacked text classification model with an ROC-AUC score of 0.95 that uses program description to predict program type in order to increase accessibility within MCMF
- Presented the data visualization product to the client representative and suggested data and research driven recommendations on improving educational equity

Function Genomics Analysis Project   *January 2023 – March 2023*

Northwestern University - Evanston, IL   https://emmayu18.github.io/Functional-Genomics-Project/

- Conducted two comparative gene expression analyses on juvenile idiopathic arthritis and IBD gene fold change data to determine whether immune system signaling pathways are over-expressed in these diseases

SKILLS

- **Technical:** Python (pandas, matplotlib, scikit-learn, spaCy, keras), R (tidyverse, tidymodels, ggplot2, shiny), Jupyter Notebook, SQL, NLP, CSS, Git, Tableau, AWS, Microsoft Office
- **Language:** English (native), Korean (fluent)